JOURNAL OF MULTIVARIATE ANALYSIS 11, 50–57 (1981)

Decreasing in Transposition Property of Overlapping Sums, and Applications

M. HOLLANDER,* F. PROSCHAN,* AND J. SETHURAMAN†

Florida State University, Tallahassee, Florida 32306

Communicated by A. Cohen

In this paper a large class of multivariate densities and frequency functions, including the multivariate Poisson distribution and the compound multivariate Poisson distribution, are shown to have the decreasing in transposition property introduced by Hollander, Proschan, and Sethuraman (1977, *Ann. Statist.* 5, 722–733). Sample applications in ecology and reliability are given; other applications to cumulation of damage and component down times are mentioned, but are not presented in detail.

1. INTRODUCTION AND PRELIMINARIES

In this paper we derive a basic theorem (Section 2, Theorem 2.1) showing that a large class of multivariate probability densities and frequency functions possess the decreasing in transposition (DT) property (Definition 1.3) introduced by Hollander, Proschan, and Sethuraman (HPS) [4]. Probability densities possessing the DT property are of interest and have application in rank order problems, as shown in detail in [4]. Of particular interest are the multivariate Poisson distribution and the compound multivariate Poisson distribution treated in Section 3. In Section 3 we present two applications, one in ecology and one in reliability. Other applications to cummulation of damage and component down times exist, but are not explicitly presented. In the remainder of this section we give some preliminaries, including definitions and theorems, which will be useful in the sequel.

Received January 4, 1979; revised March 29, 1980.

AMS 1970 subject classifications: Primary 62E99, 62H99; Secondary 62N05.

Key words and phrases: Functions decreasing in transposition, multivariate distributions, reliability, shock models.

* Research supported by the Air Force Office of Scientific Research, AFSC, USAF, under Grant No. 78-3678.

† Research supported by the United States Army Research Office, Durham, under Grant No. DAAG-29-79-C-0158.

0047-259X/81/010050–08$02.00/0

HPS define a partial ordering on n-dimensional Euclidean space (R^n) presented in Definition 1.2 below.

DEFINITION 1.1. A vector $\mathbf{x} = (x_1, ..., x_n)$ is said to be a *simple* transposition of a vector $\mathbf{x}'$ if $\mathbf{x}$ and $\mathbf{x}'$ agree in all but two coordinates, say i and j, $i < j$, $x_i < x_j$, $x'_i = x_j$, and $x'_j = x_i$; we write $\mathbf{x} >^t \mathbf{x}'$.

Thus $\mathbf{x}'$ is obtained from $\mathbf{x}$ by performing an inversion of a single pair of coordinates that occur in their natural order in $\mathbf{x}$.

DEFINITION 1.2. Let $\mathbf{x}$ and $\mathbf{x}'$ be two n-dimensional vectors such that there exists a finite number of vectors $\mathbf{x}^0, \mathbf{x}^1, ..., \mathbf{x}^k$ in R^n satisfying $\mathbf{x} = \mathbf{x}^0 >^t \mathbf{x}^1 >^t \cdots >^t \mathbf{x}^k = \mathbf{x}'$; i.e., $\mathbf{x}'$ is obtained from $\mathbf{x}$ by a finite number of simple transpositions. We say that x' is a *transposition* of $\mathbf{x}$.

HPS then define a class of functions as follows:

DEFINITION 1.3. Let $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)$, where $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ are n ordered values in R^1. We say that $f(\boldsymbol{\lambda}, \mathbf{x})$ is *decreasing in transposition* (DT) if:

(a) $f(\boldsymbol{\lambda}^\pi, \mathbf{x}^\pi) = f(\boldsymbol{\lambda}, \mathbf{x})$ for each permutation $\pi = (\pi_1, ..., \pi_n)$ of the indices $1, 2, ..., n$, where $\boldsymbol{\lambda}^\pi = (\lambda_{\pi_1}, ..., \lambda_{\pi_n})$ and $\mathbf{x}^\pi = (x_{\pi_1}, ..., x_{\pi_n})$.

(b) $\mathbf{x} >^t \mathbf{x}'$ implies that $f(\boldsymbol{\lambda}, \mathbf{x}) \geqslant f(\boldsymbol{\lambda}, \mathbf{x}')$.

The familiar concepts of majorization and Schur functions will be used in this paper, and for completeness we recall their definitions.

DEFINITION 1.4. Let $x_{[1]} \geqslant \cdots \geqslant x_{[n]}$ be a decreasing rearrangement of the coordinates of the vector $\mathbf{x}$. Let $\mathbf{x}$ and $\mathbf{y}$ satisfy

$$\sum_{i=1}^{j} x_{[i]} \geqslant \sum_{i=1}^{j} y_{[i]}, \qquad j = 1, ..., n-1,$$

$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}.$$

Then $\mathbf{x}$ is said to *majorize* $\mathbf{y}$ (we write $\mathbf{x} \geqslant^m \mathbf{y}$).

DEFINITION 1.5. A function f from R^n into R^1 is said to be *Schur-convex* (*Schur-concave*) if $\mathbf{x} \geqslant^m \mathbf{y}$ implies $f(\mathbf{x}) \geqslant (\leqslant) f(\mathbf{y})$.

HPS derive basic properties of DT functions and show their relationship to other classes of functions. They show [4, Lemma 2.2] that when $f(\boldsymbol{\lambda}, \mathbf{x})$ is of the form $g(\boldsymbol{\lambda} - \mathbf{x})$, $f(\boldsymbol{\lambda}, \mathbf{x})$ is a DT function if and only if $g(\mathbf{y})$ is a Schur-concave function. They also show [4, Theorem 3.2] that the DT property is preserved under positive mixtures and [4, Theorem 3.6] that the DT property is preserved under products of positive DT functions. They also prove a "composition" theorem for DT functions [4, Theorem 3.3] and a "preservation" theorem [4, Theorem 3.7] for Schur-concave functions under an integral transform where the kernel is DT.

By restricting $g_2(\mathbf{y}, \mathbf{z})$ in the HPS "composition" theorem to be of the form $g(\mathbf{y} - \mathbf{z})$, where $g(\mathbf{w})$ is Schur-concave, we have the following.

THEOREM 1.6. *Let $g_1(\mathbf{x}, \mathbf{y})$ be a DT function and $g(\mathbf{w})$ be a Schur-concave function such that*

$$f(\mathbf{x}, \mathbf{z}) = \int g_1(\mathbf{x}, \mathbf{y})\, g(\mathbf{y} - \mathbf{z})\, d\mu(\mathbf{y})$$

is well defined, where μ is a positive permutation invariant measure. Then $f(\mathbf{x}, \mathbf{z})$ is a DT function.

Theorem 1.6 generalizes the following result of Marshall and Olkin [7, Theorem 2.1].

THEOREM 1.7 (Marshall and Olkin [7]). *Let $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ be Schur-concave functions such that*

$$f(\mathbf{x}) = \int g_1(\mathbf{x} - \mathbf{y})\, g_2(\mathbf{y})\, d\mu(\mathbf{y})$$

is well defined, where μ is as in Theorem 1.6. Then $f(\mathbf{x})$ is a Schur-concave function.

2. DT PROPERTY OF OVERLAPPING SUMS

In this section we state and prove our main theorem. For $k = 2, 3, \ldots, n$, let

$$\begin{aligned} I_k &= \{I\colon I \text{ is a subset of size } k \text{ from } \{1, \ldots, n\}\}, \\ I_{k,i} &= \{I \in I_k\colon i \in I\}. \end{aligned} \tag{2.1}$$

THEOREM 2.1. *Let $\mathbf{X} = (X_1, \ldots, X_n)$, $\{X_I, I \in I_k\}$, $k = 2, \ldots, n$, be independent collections of random variables. Let $\mathbf{X}$ have a DT density function. Let the random variables in $\{X_I, I \in I_k\}$ be i.i.d. and have a common log-concave density g_k, $k = 2, \ldots, n$. For $i = 1, \ldots, n$, let*

$$Z_i = X_i + \sum_{\substack{I \in I_{k,i} \\ k \geq 2}} X_I. \tag{2.2}$$

Then $\mathbf{Z} = (Z_1, \ldots, Z_n)$ has a DT density function.

Remark 2.2. Note that the summands in the $Z_1,\ldots,Z_n$ overlap considerably. For example, X_{12} appears in the expressions for Z_1 and Z_2, X_{123} appears in the expressions for Z_1, Z_2, and Z_3, etc. Thus the inheritance of the DT property of $\mathbf{Z}$ from that of $\mathbf{X}$ is complicated by the overlapping of the X_I's.

Remark 2.3. Theorem 2.1 takes on added interest if we note that the tempting conjecture that "the convolution of DT functions is DT" is *false*. An even more tempting conjecture that "the convolution of a DT density and an exchangeable density is DT" is also false. These facts point up the need for the log-concavity of the g_k density in the statement of Theorem 2.1.

Remark 2.4. Random variables $Z_1,\ldots,Z_n$ of the type specified in (2.2) arise routinely in shock models, inventory problems, biometric models, and elsewhere in multivariate statistics, where the occurrence of an event *simultaneously* affects two or more random variables of interest. A classical example is the multivariate exponential of Marshall and Olkin [6], where a shock of type $(i_1, i_2,\ldots, i_k)$ results in the *simultaneous* failure of components $i_1,\ldots, i_k$. In Section 3, we give illustrative examples from reliability theory, in particular, in which the main theorem would apply.

To prove the main result, we shall find it helpful to have available the following lemma:

LEMMA 2.5. *For some $k \geqslant 2$, let $\{X_I, I \in I_k\}$ be i.i.d. random variables with a common log-concave density function g (with respect to the counting measure on a lattice or the Lebesgue measure). Let*

$$W_i = \sum_{I \in I_{k,i}} X_I, \qquad i = 1,\ldots, n. \tag{2.3}$$

Let $f(w_1,\ldots, w_n)$ be the density function of $\mathbf{W} = (W_1,\ldots, W_n)$. Then f is a Schur-concave function.

Proof. We will now give a proof for the case where g and f are density functions with respect to counting measures. The proof for the case where g and f are density functions with respect to Lebesgue measures is similar and will be omitted.

Notice first that $W_1,\ldots, W_n$ are exchangeable and hence $f(w_1,\ldots, w_n)$ is permutation invariant. Fix $w_3,\ldots, w_n$ and define

$$A_{w_1,w_2} = \{W_1 = w_1, W_2 = w_2, W_3 = w_3,\ldots, W_n = w_n\}.$$

To show that f is Schur-concave, in view of a result of Hardy, Littlewood, and Pólya [2, p. 47] that $(w'_1,\ldots, w'_n)$ can be derived from $(w_1,\ldots, w_n)$ by a finite number of pairwise averages of components when $w \geqslant m_{w'}$, it suffices to show that

$$P(A_{w_1,w_2}) \leqslant P(A_{w'_1,w'_2}) \tag{2.4}$$

whenever

$$(w_1, w_2) \geqslant^m (w_1', w_2'),$$

i.e., whenever

$$0 \leqslant \beta \leqslant \alpha \leqslant w^*, \tag{2.5}$$

where

$$w_1 = w^* + \alpha, \qquad w_2 = w^* - \alpha, \qquad w_1' = w^* + \beta, \qquad w_2' = w^* - \beta.$$

The following classes of subsets of $\{1,..., n\}$ will help us to divide the elements of I into those that contain both 1 and 2, those that do not contain either 1 or 2, and those that contain exactly one of 1 and 2. Define

$$\mathscr{V} = \{V: V \text{ is a subset of size } (k-2) \text{ from } \{3,..., n\}\},$$

$$\mathscr{U} = \{U: U \text{ is a subset of size } (k-1) \text{ from } \{3,..., n\}\},$$

and

$$\mathscr{S} = \{S: S \text{ is a subset of size } k \text{ from } \{3,..., n\}\}.$$

Any I in I_k is of the form $12V$, $1U$, $2U$, or S. Fix numbers x_{12V}, $V \in \mathscr{V}$, x_S, $S \in \mathscr{S}$, and x_U, $U \in \mathscr{U}$, and define the event B as follows:

$$B = \{X_{12V} = x_{12V}, V \in \mathscr{V}, X_S = x_S, S \in \mathscr{S}, X_{1U} + X_{2U} = 2x_U, U \in \mathscr{U}\}. \tag{2.6}$$

Let

$$2Y_U = X_{1U} - X_{2U}, \qquad U \in \mathscr{U}.$$

Then the density function of $\{Y_U, U \in \mathscr{U}\}$ conditional on B is

$$\frac{\prod_U g(x_U + y_U)\, g(x_U - y_U)}{\sum_{y_U, U \in \mathscr{U}} \prod_U g(x_U + y_U)\, g(x_U - y_U)}, \tag{2.7}$$

and the conditional probability of A_{w_1, w_2} given B is

$$\begin{aligned} P(A_{w_1,w_2} \mid B) &= \frac{\sum_{\sum y_U = \alpha} \prod_U g(x_U + y_U)\, g(x_U - y_U)}{\sum_{y_U, U \in \mathscr{U}} \prod_U g(x_U + y_U)\, g(x_U - y_U)} \\ &\quad \text{if } w_j = \sum_{j \in V} x_{12V} + \sum_{U: j \in U} x_U + \sum_{S: j \in S} x_S \; j = 3, 4,..., n \text{ and } w^* = \sum_V x_{12V} + \sum_U x_U \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

Now for each x_U, $g(x_U + y_U)\, g(x_U - y_U)$ is symmetric and log-concave in y_U. The numerator in the expression for $P(A_{w_1,w_2} \mid B)$ is therefore a convolution of symmetric log-concave densities and hence from Theorem 1 of Karlin and Proschan [5] is symmetric and log-concave in α. Thus $P\{A_{w_1,w_2} \mid B\}$ is decreasing for $\alpha > 0$ and so if (2.5) is satisfied, we obtain

$$P(A_{w_1,w_2} \mid B) \leqslant P(A_{w_1',w_2'} \mid B).$$

Inequality (2.4) now follows by unconditioning, and thus we have established that f is Schur-concave. ∎

We now present the proof of the main theorem.

Proof of Theorem 2.1. Let

$$X_i^{(k)} = \sum_{I \in I_{k,i}} X_I, \qquad i = 1,\ldots, n, \quad k = 2,\ldots, n.$$

Let $\mathbf{X}^{(k)} = (X_1^{(k)},\ldots, X_n^{(k)})$, $k = 2,\ldots, n$, and $\mathbf{X} = (X_1,\ldots, X_n)$. Then

$$\mathbf{Z} = \mathbf{X} + \mathbf{X}^{(2)} + \cdots + \mathbf{X}^{(n)}.$$

From Lemma 2.5, the density functions of $\mathbf{X}^{(2)},\ldots, \mathbf{X}^{(n)}$ are all Schur-concave. Since the density function of $\mathbf{Z}$ is the convolution of a DT and several Schur-concave densities, it follows from Theorems 1.6 and 1.7 that the density function of $\mathbf{Z}$ is DT. ∎

3. Applications

In this section we present models from reliability and ecology in which the main theorem would be used to establish the DT property of probability densities of interest. The subsequent exploitation of the DT property to obtain results in rank order statistics has been described in [4] and will not be repeated here.

3.1. *Multivariate Exponential of Marshall–Olkin*

Marshall and Olkin [6] introduce the widely used multivariate exponential in which a shock of type I causes the simultaneous failure of components in I, where I is a subset of $\{1, 2,\ldots, n\}$. The shocks of type I are governed by a Poisson process with rate λ_I. The $2^n - 1$ such Poisson processes are assumed mutually independent.

Numbers of replacements, amounts of damage cumulated, and down times for components. Assume failed components are immediately replaced. Then the numbers $Z_1^{(R)},\ldots, Z_n^{(R)}$ of replacements in a fixed interval of time have a joint multivariate Poisson distribution. (See Teicher [8] and Dwass and Teicher [1].) If $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$; $\lambda_{ij} = \lambda^{(2)}$ for all pairs $i < j$; $\lambda_{ijk} = \lambda^{(3)}$ for all triplets $i < j < k$, etc., then from Theorem 2.1 we deduce that $(Z_1^{(R)},..., Z_n^{(R)})$ has a DT frequency function.

One obvious consequence is that the most likely outcome is $Z_1^{(R)} \leqslant Z_2^{(R)} \leqslant \cdots \leqslant Z_n^{(R)}$. Moreover, a partial ordering of likelihoods of outcomes is given by the DT partial ordering.

In a similar way, under the hypotheses of Theorem 2.1, we can show that vector of cumulative damages $Z_1^{(D)},..., Z_n^{(D)}$ experienced respectively by components 1,..., n, is governed by a DT density.

Similarly, the joint density of component down times is DT.

Note that although we have chosen the well-known Marshall–Olkin MVE for illustrative purposes, the DT property obtained holds for a far larger class of underlying multivariate densities and frequency functions.

3.2. *Compound Multivariate Poisson*

The compound multivariate Poisson distribution described below generalizes the compound bivariate Poisson distributions considered by Holgate [3] in the context of certain ecological situations. Let $Z_1, Z_2,..., Z_n$ denote the number of individuals of type 1, 2,..., n, respectively, in a quadrat of land. We suppose that these individuals arise from independent clusters and assume that the number of clusters N is a Poisson random variable with parameter m. In cluster j, there are

$$Z_i^j = X_i^j + \sum_{k=2} \sum_{I \in I_{k,i}} X_I$$

individuals of type i, $i = 1,..., n$, where X_i^j, X_I, $I \in I_k$, $k = 2,..., n$, are independent Poisson random variables with parameter m_i, m_I, $k = 2,..., n$, respectively, and satisfying $m_I = m_k^*$ for $I \in I_k$, $k = 2,..., n$, for each cluster j. We have already stated that the clusters are assumed to be independent. Thus $\mathbf{Z}^j = (Z_1^j,..., Z_n^j)$ has a multivariate Poisson distribution. The vector $\mathbf{Z}$ of the number of individuals of the different types in the quadrat is given by

$$Z = \mathbf{Z}^1 + \cdots + \mathbf{Z}^N$$

and is said to have a compound multivariate Poisson distribution. We now show that the density of $\mathbf{Z}$ is DT. Conditional on $N = N_0$, $\mathbf{Z}$ is the sum of N_0 i.i.d. multivariate Poisson vectors and therefore is multivariate Poisson from the closure under convolution property of the multivariate Poisson established by Dwass and Teicher [1]. Thus the conditional density of $\mathbf{Z}$ is also DT. This proof also shows that the random number of clusters N could be any random variable taking values on the positive integers.

Acknowledgments

We are grateful to the referees and Associate Editor for many helpful suggestions.

References

[1] Dwass, M., and Teicher, H. (1957). On infinitely divisible random vectors. *Ann. Math. Statist.* **28** 461–470.

[2] Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952). "*Inequalities*." Cambridge Univ. Press, London/New York.

[3] Holgate, P. (1966). Bivariate generalizations of Neyman's Type *A* distribution. *Biometrika* **53** 241–245.

[4] Hollander, M., Proschan, F., and Sethuraman, J. (1977). Functions decreasing in transposition and their applications in ranking problems. *Ann. Statist.* **5** 722–733.

[5] Karlin, S., and Proschan, F. (1960). Pólya type distributions of convolutions. *Ann. Math. Statist.* **31** 721–726.

[6] Marshall, A. W., and Olkin, I. (1967). A multivariate exponential distribution. *J. Amer. Statist. Assoc.* **62** 30–44.

[7] Marshall, A. W., and Olkin, I. (1974). Majorization in multivariate distributions. *Ann. Statist.* **2** 1189–1200.

[8] Teicher, H. (1954). On the multivariate Poisson distribution. *Skand. Aktuarietidskr.* **37** 1–9.